DOCUMENT RESUME

ED 382 677 TM 023 124

AUTHOR Mitchell, Ruth
TITLE The Promise of Performance Assessments: How To Use Backlash Constructively.
PUB DATE Apr 95
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cognitive Processes; *Constructivism (Learning); Criterion Referenced Tests; *Curriculum Development; *Educational Assessment; Educational Change; Elementary Secondary Education; *Standards; Test Construction; *Test Use
IDENTIFIERS *Backlashes; *Performance Based Evaluation

ABSTRACT

This is an age of backlash, as illustrated by many developments in the field of education. Backlash against performance assessment is very evident, but it must be remembered that backlash has its uses. It forces advocates to rethink, reformulate, and restate why they put so much faith in the program under attack. Events have pushed advocates to sharpen the case for the efficacy of performance assessment in educational reform. Performance assessment is defined as evaluation of educational progress that is standards or criterion referenced, and which requires direct demonstration of knowledge and skill. Educational factors that resist backlash include the recognition that performance assessment is a reform strategy intended to change curriculum and instruction as well as assessment. It also should be recognized that performance assessment is based on the constructivist theory of cognition and that it has been bolstered by the standards movement. The existence of the backlash is a testimony to effectiveness, because new assessments clearly threaten the status quo. (SLD)

The Promise of Performance Assessments: How to Use Backlash Constructively

Ruth Mitchell, American Association for Higher Education, Washington DC/ Pelavin Associates Senior Consultant

This is an age of backlash, against the women's movement, against affirmative action, and as Lynn Olson makes clear in *Education Week*, against performance assessment.[1] In California, as we all know, the California Learning Assessment System (CLAS), the successor to the California Assessment Program (CAP), died last fall, buried under an avalanche of misinformation about the purposes and the content of the assessments. The governor would like to replace CLAS with "traditional" assessments which provide individual student scores.

In an article describing the political and educational reasons for opposition, Lynn Olson provides us with many more examples of the backlash against performance assessment. But backlash has its uses. It forces advocates to rethink, reformulate, and restate why they put so much faith in the program under attack. Events have pushed them to sharpen their case for the efficacy of performance assessment in educational reform.

Let's just for the record be clear about what we're talking about: performance assessment is evaluation of educational progress which is standards or criterion referenced, and requires direct demonstration of knowledge and skill. It isn't multiple-choice, norm-referenced or machine-scorable. The term "performance assessment" includes constructed responses of all kinds and various lengths; open-ended questions; portfolios; exhibitions in the sense used by the Coalition of Essential Schools; interviews; observational records; written, spoken, and videotaped responses. To my mind, there is no useful distinction among

[1] *Education Week*. XIV.26, (March 22, 1995), 1 and 10

the terms used for this kind of assessment, such as "authentic," "alternative," "performance-based," or "non-traditional." They all imply active student production of evidence of learning—not multiple-choice, which is essentially passive selection among preconstructed answers.

Performance assessment as reform strategy

When performance assessment began to be used widely in the late 1980's, its most attractive feature was its potential for reforming classroom practice. Since teaching to the test is inevitable—it is only human nature to perform for a goal—then the quality of the test would determine the quality of teaching. If the assessment required production and application instead of memorization and recognition, then teaching would also have to include writing, reasoning, and the demonstration of understanding.

The intention was not to promote measurement-driven instruction, but to use a change in assessment as a way into the system. Education in K-12 schools has developed as a closed system with a hard shell which resists outside influences. To get its attention, you have to tap into the system through its few vulnerable spots. External accountability is the softest of these spots and accountability depends on assessment. So reformers seized on a change in assessment as a way to get into the system and bring about changes. But they did not intend to reform only accountability, because most assessment actually takes place at the classroom level.

Dan Koretz was quoted by Lynn Olson as saying "There was an initial period of enormous enthusiasm which, in my judgment, was often unrealistic...and now people are going to have to start asking: Are we getting what we're paying for?" Among others, I personally was criticized for painting a rosy picture of assessment-driven instruction. Critics assumed that reformers had a naive view of educational reform: if you changed assessments, everything else would miraculously follow. Reformers did not believe that. They believed that changes in assessment were both a starting point and a *sine qua non* for changes in teaching and learning.

Under the scrutiny intensified by backlash, that belief remains and is strengthened. Assessment must focus on valued aspects of learning if they are to be taught. The forms of assessment matter, despite arguments that multiple-choice can assess thinking skills just as well as constructed responses. Take writing as an example: because it is now widely recognized that the quality of writing cannot be judged by multiple-choice items, the amount of writing in American classrooms has increased and "the writing process" is almost universally known, if not understood or taught well. The amount of writing in classrooms has increased only because "writing samples" are required to assess writing: states and districts which cling to mutiple-choice usually boast a "writing sample." Since assessments such as the Vermont portfolio require written explanations and applications in mathematics, the same process is slowly beginning in mathematics.

When reformers placed their faith in assessment as a way into the system, they emphasized (although apparently not enough, considering how the message was misunderstood) that other components of the system would have to change if the promise of performance assessment was to be realized. These components include: professional development; goals, standards, and expectations; curriculum; pedagogy; textbooks and materials; preservice teacher education; public understanding of the purposes and practices of education; the distribution of funds; state and national legislation, particularly with regard to Chapter 1. Changing assessments ultimately affects all of these. In the case of Chapter 1, now Title 1 under reauthorization, a change in assessment requirements has removed a hoary old argument for retaining nationally published norm-referenced tests. We used to hear school administrators say that they might as well keep these tests since they had to use them for Chapter 1 children. No longer. (An interesting subject for research would be to find out how many SEAs and LEAs know this—and are acting on it.)

Reformers maintained then and maintain now that taking part in assessment, as designers, administrators, and scorers, is among the best kinds of professional development available for teachers—indeed for all school personnel. Ample anecdotal evidence supports teachers who claim that participating . . scoring sessions opened their eyes to the need for

revision of their aims and methods. Richard Hill tells of a group of English teachers in Kentucky who participated in that state's portfolio assessment and whose scores were consistently altered downwards on review; finally, when the teachers rescored in company with their reviewers, they found that they had been rewarding "correct" writing (i.e. accurate grammar, spelling, and punctuation) but had ignored the quality of content. The imperative for change was clear to them.

Equity issues

However, changing the assessments doesn't necessarily bring with it all the systemic reforms that are necessary. (The other presentations in this session will make that abundantly clear.) A conference held two years ago in Washington attempted to focus on the equity issues involved in moving from multiple-choice norm-referenced tests to performance assessments. Quite correctly, minorities are suspicious of forms of assessment that seem to bring back the teacher judgments under which minority students suffered discrimination. Paper after paper at the conference made arguments which expressed apprehension (because not much research-based evidence was available) about the adverse effects of performance assessment on minority students; in fact, these writers were indicting the quality of teaching, the instructional materials, the opportunities to learn for all students, and not the assessments as such.[2] By implication they were making the same point as was made in advocating performance assessment: it is a reform strategy, not a simple replacement for traditional tests. It won't work for any students if they are faced with having to apply and explain a mathematical principle when they've only been taught simple algorithms.

Standards and assessment

The case for performance assessment has been enormously strengthened by the burgeoning of the standards movement. In the case of Title 1, states and districts must hold all children to the same high

[2] The papers from the conference have now been published in *Equity and Excellence in Educational Testing and Assessment*, edited by Michael T. and Arie L. Nettles; Boston/Dordrecht/London:Kluwer Academic Publishers, 1995

standards, and to get Goals 2000 money, states must establish standards or adopt national or state standards of equal rigor. Since I have been involved with several cities in the writing of content standards, I am now amazed at myself—I wrote a book on performance assessment in 1991 without explicitly mentioning standards. They were implicit, of course, as they are in the practice of good teachers; since group grading began, standards have been expressed as rubrics.

Now that three kinds of standards—content, performance, and opportunity to learn (OTL)—have been clearly delineated, the frame of the puzzle is in place, and the place of assessment in each case becomes clear. Content standards are statements about what is to be assessed and performance standards describe levels of achievement across a domain (a kind of grand rubric). Standards are written with assessable verbs which themselves demand performance assessment: "apply problemsolving strategies to...construct tables, charts, and graphs to summarize data...design a statistical survey...analyze characteristics of ...describe the nature and role of national, state, and local government."

Opportunity to learn standards will list those attributes of school context which enable access for all students to content standards. Contexts, however, can vary tremendously: learning can take place in situations where it might seem most unlikely and vice versa. Performance assessments, particularly portfolios, can provide information about opportunities to learn that would be unavailable in any other way. A set of classroom or schoolwide portfolios can tell an observer what topics students were introduced to and what they were asked to do; to what depth and in what variety; and how learning opportunities vary in different classrooms. (The example from Kentucky cited above makes this point dramatically: the portfolios demonstrated a restricted opportunity to learn in language arts classrooms which were focused on correctness, not content.) Such information is of course after the fact, which argues strongly for not attaching high stakes decisions to performance asessments until there is more research-based information about OTL. The point here is that portfolios and the examination of student work involved in performance assessments are essential to establishing OTL standards.

The question of assuring equitable opportunity to learn as well as improving the quality of student learning forces a focus on the two major failures of systemic reform: inadequate professional development and miserable public relations. You will hear over and over again in the reports of our case studies about the consequences of inadequate professional development for teachers. It is manifested in the poor quality of the assessments they design; in the rubrics which are sometimes no more than checklists; in the unchanged classroom behavior. When you hear teachers and administrators complain that performance assessment takes too much time away from instruction, you know they haven't got it. Only professional development will enable them to understand that assessment and instruction should be seamless.

Public information and misinformation

If anything, lack of explanation to the public is even more damaging to performance assessment as a reform strategy. I referred earlier to the closed educational system: reformers are now reaping the bitter harvest of not communicating their reasons for changes to parents, business, and legislators. The public does not understand why schools should be different from the ones they attended. They do not buy the argument that economic competitiveness depends on all students exercising higher-order thinking skills.[3] They want safe and orderly schools, and high standards, but "the basics" are essential in their minds.[4] Educators have not explained clearly the shift from behaviorism to constructivism in education, so that parents, legislators, and test publishers still believe in the acquisition of knowledge by little bits, which may or may not add up to a concept. Jean Johnson and John Immerwahr are cautious in *First Things First*: "...leaders may decide that the public's point of view (in whole or in part) is mistaken...this is warranted if, after honest self-scrutiny, leaders are convinced their approach—not the public's—will truly help children and their families" (p. 39).

---

[3] *CrossTalk: The Public, the Experts, and Competitiveness.* A Research Report from The Business-Higher Education Forum and The Public Agenda Foundation, February 1991.
[4] *First Things First. What Americans Expect from the Public Schools.* A Report from Public Agenda, 1994

Public ignorance leads directly to demands for easily understood numbers and familiar grades and for individual student scores from state-level tests. Trying to please an uninformed public and also move the system in the right direction, states and districts have kept their multiple-choice norm-referenced tests alongside the performance assessments, thus confusing everyone.

However, an increasing proportion of public opposition is not the result of missing information but of a deliberate campaign of misinformation mounted by organized groups, with respectable-sounding names like Capitol Resource Institute, the Claremont Institute, and the United States Justice Foundation in Californa, and the Rutherford Institute in Virginia. Such groups have been particularly effective in Virginia, Pennsylvania, Georgia, and California, but they are a threat in all states.[5] Their agenda has gone beyond attacks on performance assessment in California and outcomes-based education elsewhere to a concerted attack on teaching higher-order thinking skills: they believe that the business of schools should be confined to what can be tested by multiple-choice—depersonalized and decontextualized knowledge.

Although membership in this groups is tiny, their influence is magnified precisely because of public ignorance. They supply misinformation which fills a vacuum in public understanding. During the 1994 furor in California over the content of the writing/reading assessment, these groups distributed so-called "state assessments" which did not originate with CLAS at all; since they had no reliable authentic information, the public tended to believe the groups' propaganda. For example, the Capitol Resource Institute distributed *Examples of CLAS*, which began: "The California Department of Education continues to violate state law and parents' rights. Capitol Resource Institute is revealing 'secret' material from the California Learning Assessment System, in order to expose this violation." They follow this introduction with criticisms of stories which they maintain formed part of the CLAS assessment despite repeated denials from the California State Department of Education.

---

[5]In the *Education Week* article, Lynn Olson refers to the Claremont Institute's attack on the California Curriculum Frameworks, which are regarded as models nationwide.

Because of the recent change of leadership in the Congress and the increased influence of organized opposition groups, test publishers have regained their privileged access in state departments of education and on Capitol Hill. In the corridors of Congress, the test publishers' lobbyists present themselves as businesspeople talking to other businesspeople. They plead for the test-publishing business because it is successful on many levels, including employment. Multiple-choice, norm-referenced, machine-scorable tests appear to have a sound track record and they don't mess with thinking. The test publishers' case is bolstered by the organized opposition, who have allies such as House Majority Leader Richard Armey. In a letter to his colleagues opposing Goals 2000, Armey wrote: "Soon, children are taking tests with open-ended questions like: 'Three things I don't like about my parents are...' Any wonder why it angers?"[6]

Refuting such nonsense with logic is futile. Instead, reformers must adopt a policy of openness and inclusion towards parents and the community, as they have in places where the educational community and the public are actively cooperating, such as Forth Worth Independent School District in Texas. There, performance assessment is accepted as the corollary to applied learning, but applied learning schools are a choice and the public is not forced to accept unfamiliar school practices.

Developments at the higher education level

Some familiar practices are changing, however, at the higher education level[7]. As we all know, reform at the high school level is harder than at elementary and middle levels, largely because high school curriculum and instruction is driven by college and university admissions. The facade presented by higher education is cracking. Six state university systems have grants from the Pew Charitable Trusts to

[6] Letter to Congressional colleagues dated 6 October 1993.

[7] With at least one notable exception, however. Higher education institutions in Kentucky are unhappy with the predictive quality of scores derived from the Kentucky Instructional Results Information System (KIRIS), apparently because they misunderstand the nature of KIRIS, which is intended to produce accountability numbers at the school, not the individual student, level.

study admissions based on student portfolios or other classroom achievements; the interest is such that a recent conference on these admissions policies attracted other state university systems who also want to experiment with what is called "proficiency-based admissions," an unfortunate name. At the same time, the number of institutions, particularly small liberal arts colleges, which do not require the SAT and which will accept portfolios for admission is growing—about 200 public and private institutions by the latest FairTest count. Michael Kirst of Stanford University has publicly called for the replacement of the SAT with achievement tests, using the argument on which reformers of assessment rely—that testing achievement rather than aptitude will reinforce curriculum and instruction in high schools.[8]

Classroom assessment and accountability

Please note that proposed "proficiency-based admissions" relies on classroom assessment. The relationship of classroom assessment to accountability assessment is an ongoing puzzle. The assessment-reform movement started off confident that the two could be accomplished with a single instrument: classroom portfolios could be sampled at the district, regional, and state level in a pyramid of review for accountability, it seemed. But the work of Dan Koretz and the RAND team in Vermont questioned that confidence. A recent *Evaluation Comment* from CRESST asks *Whose Work Is It?*[9] The authors are referring not only to the obvious question of adult assistance with assignments if they are taken home or even discussed at home, but also to the much subtler consequences of believing that learning takes place in a social context, a fundamental tenet of constructivism. If a student's portfolio is influenced by peers commenting on rough drafts, by the teacher's support, (which may differ according to perceived need) or by class

[8] Michael Kirst and Henry Rowen: "Scrap the SATs for Achievement Tests," *The Washington Post*, Friday September 16, 1994, page A27. We note, as Kirst and Rowen do not, that College Board Achievement Tests (called SAT II) are still mostly multiple-choice in form, with the exception of a "writing sample." To achieve the effect on curriculum that Kirst and Rowen hope for, achievement would have to be measured by performance assessment, probably portfolios.

[9] Maryl Gearhart and Joan L. Herman, *Portfolio Assessment: Whose Work Is It? Issues in the Use of Classoom Assignments for Accountability.* CRESST, 1995

discussions, can it be said to be the student's own work? Gearhart and Herman suggest a number of strategies for evaluating the student's own work from a portfolio, but they also ask: "But *can* portfolio assessment provide us with valid indices of student competencies usable for large-scale accountability?" The question remains unanswered.

The obvious next question is: why are we concerned with large-scale accountability? Rethinking the purposes of assessment, performance assessment advocates maintain that the primary purpose of any assessment must be to improve teaching and learning.[10] Assessment provides feedback to students and teachers in order for them to readjust their activities in reaching content standards or goals. Accountability is a secondary—and essentially political—purpose for assessment.

That assessment is primarily instructional feedback is a challenging notion to classical psychometricians, whose professional sights are fixed on perfecting large-scale acccountability methods. The difference in points of view is sharply brought into focus by Gearhart and Herman's aticle. They acknowledge the value of portfolios as examples of good classroom instruction, which "according to current pedagogical and curriculum reforms involves an engaged community of practititioners in a supportive learning process" (p. 3). But psychometrics is stymied by a product contaminated in its view by assistance from others. The performance asessments advocates assumed that performance assessment would prompt research and rethinking by the psychometric community. Gearhart and Herman point out that portfolios replicate "what 'real' writing entails, in that writing is often a very social endeavor" (p. 3). Since cooperation, collaboration, and collective work are valued, why is it important to assess an individual's achievement? If schools and educators are being urged to reduce the gap between the classroom and the real world, then the technical quality of measurement in the classroom should also be

[10]The Nation Forum on Assessment (NFA) has a statement of criteria for assessment which places this principal first. (The criteria are reprinted in *Equity and Excellence in Educational Testing and Assessment* [see footnote 2], pp.150-3.) The NFA is currently circulating for comment a draft of a document in which the criteria are elaborated as standards with indicators for each standard.

reconsidered. In the "real" world, assessment is instantaneous; temporary (connected to an immediate purpose or need); negotiated; and lacking in reliability, although its validity may be high. Real-world assessment would give a psychometrician hives.

The field needs an extended psychometrics which would embrace collective products, as well as evidence of student achievement which takes different forms to show progress to the same goal or standard. The psychometric community is responding, but it needs to ask questions about the purposes of its rigor. Do we really need the statistical apparatus that justifies a profession? Perhaps significantly, school people seem less troubled by psychometric imperfections than are the professionals. We thought it possible that psychometric immaturity would be used as an argument against adopting performance assessments in some schools, but that does not seem to be the case from our studies.[11]

Summing up

Performance assessment like all educational issues is affected partly by educational and partly by political pressures. It may be helpful to summarize the situation in those terms.

Here are the educational factors that resist backlash and may indeed be strengthened by it:

- performance assessment is a reform strategy intended to change curriculum and instruction as well as assessment;
- performance assessment is based on the constructivist theory of cognition;
- performance assessment has been bolstered by the standards movement, which makes a spiral of standards, curriculum, assessments, and professional development, each feeding and modifying the others;

[11]We find that teachers echo experts' criticisms of state-level assessments. They apparently worry that their work will be judged by state-level assessments which are not reliable or valid, since state-level assessments are published in newspapers. But teachers by and large do not criticize classroom-level assessments on psychometric grounds.

- psychometrics is moving in the direction of suppporting performance assessment, but is not yet meeting new needs;
- Classroom assessment is the major focus, with accountability as secondary consideration.

These are political factors, mostly negative in their effect:

- performance assessment and standards-driven reform both depend on systemic reform throughout the school system;
- the reform has not so far maintained an even pace, so that professional development is not yet supporting standards and assessment adequately;
- public information and education has been neglected by the educational system, which has reacted against performance assessments in, for example, California, Arizona, Georgia, and Littleton CO;
- there is widespread misunderstanding of performance assessment, its aims, methods, and strategies, in both the education and the wider community;
- there is also a deliberate organized campaign against educational reform, which rejects standards, performance assessment, and conceptual teaching;
- In consequence of poor public understanding and the misinformation widely distributed by organized opposition groups, additional money for schools and educational reform does not flow.

In sum, the backlash against performance assessment has sharpened advocacy and highlighted problems. The existence of backlash is testimony to effectiveness—new assessments obviously threaten the status quo. Performance assessment is here to stay, but it is now in the stage beyond an exciting idea with potential. It needs to become leaner and meaner, less fuzzy and more focused. Above all, it needs to become a routine component of an educational system directed entirely towards student success.